**Challenges Faced Report**

**1. Introduction**

This section details the challenges encountered during the flight fare prediction project and the strategies used to overcome them.

**2. Data Preprocessing Challenges**

* **Missing Values**:  
  Missing values in important columns such as Arrival\_Time and Dep\_Time were problematic. Simply dropping these rows would have resulted in a significant loss of data.

**Solution**: Missing values were handled through imputation using mean and median methods, ensuring minimal data loss.

* **Categorical Feature Encoding**:  
  The dataset contained several categorical features (e.g., Airline, Source, Destination) that needed transformation into numerical values. One-hot encoding increased the dataset’s dimensionality, which could potentially lead to overfitting.

**Solution**: One-hot encoding was used judiciously, followed by feature selection techniques to reduce dimensionality.

**3. Model Training and Tuning Challenges**

* **Hyperparameter Tuning**:  
  Finding the right set of hyperparameters for complex models like XGBoost and Gradient Boosting required considerable computational resources and time.

**Solution**: Grid search and cross-validation were employed to optimize hyperparameters. Early stopping was used to prevent overfitting during model training.

* **Overfitting**:  
  Powerful models like XGBoost tend to overfit, especially on small datasets.

**Solution**: Cross-validation and regularization techniques (like L2 in XGBoost) were used to mitigate overfitting.

**4. Computational Resource Constraints**

* **Long Training Times**:  
  Models like XGBoost and Gradient Boosting required substantial training time due to the complexity of the algorithms and the number of hyperparameters to tune.

**Solution**: Training was parallelized where possible, and simpler models like Random Forest were used for initial experimentation.

* **Model Deployment**:  
  For real-time applications, the model’s complexity (especially XGBoost) could introduce latency issues during prediction.

**Solution**: Simplified versions of the model (using fewer trees or smaller depths) were considered for real-time deployment, with periodic retraining to maintain accuracy.

**5. Model Interpretability**

* **Complexity of Boosting Models**:  
  Although models like XGBoost are highly accurate, they are less interpretable compared to simpler models like Linear Regression or Decision Trees.

**Solution**: SHAP (SHapley Additive exPlanations) was explored to provide insights into the model’s predictions, allowing for more transparency in decision-making.

**6. Conclusion**

The flight fare prediction project faced several challenges related to data preprocessing, model training, and resource constraints. However, through careful imputation, hyperparameter tuning, and model selection, the project succeeded in developing a highly accurate model (XGBoost). Future efforts should focus on improving model interpretability and optimizing deployment for real-time applications.